Fundamentals of Data Science

The F test

Ramesh Johari

The recipe

The hypothesis testing recipe

In this lecture we repeatedly apply the following approach.

- *If* the true parameter was θ_0, *then* the test statistic $T(\mathbf{Y})$ should look like it would when the data comes from $f(Y|\theta_0)$.
- We compare the *observed* test statistic T_{obs} to the *sampling distribution under* θ_0.
- If the observed T_{obs} is unlikely under the sampling distribution given θ_0, we *reject the null hypothesis that* $\theta = \theta_0$.

The theory of hypothesis testing relies on finding *test statistics* $T(\mathbf{Y})$ for which this procedure yields as high a power as possible, given a particular size.

The F test in linear regression

Multiple regression coefficients

We again assume we are in the linear normal model:
$Y_i = \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i$ with i.i.d. $\mathcal{N}(0, \sigma^2)$ errors ε_i.

The Wald (or t) test lets us test whether *one* regression coefficient is zero.

Multiple regression coefficients

We again assume we are in the linear normal model:
$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i$ with i.i.d. $\mathcal{N}(0, \sigma^2)$ errors ε_i.

The Wald (or t) test lets us test whether *one* regression coefficient is zero.

What if we want to know if *multiple* regression coefficients are zero? This is equivalent to asking: *would a simpler model suffice?* For this purpose we use the F test.

The F statistic

Suppose we have fit a linear regression model using p covariates. We want to test the null hypothesis that *all* of the coefficients $\beta_j, j \in S$ (for some subset S) are zero.

Notation:

- $\hat{\mathbf{r}}$: the residuals from the full regression ("unrestricted")
- $\hat{\mathbf{r}}^{(S)}$: the residuals from the regression *excluding* variables in S ("restricted")

The F statistic is:

$$F = \frac{(\|\hat{\mathbf{r}}^{(S)}\|^2 - \|\hat{\mathbf{r}}\|^2)/(p - |S|)}{\|\hat{\mathbf{r}}\|^2/(n-p)}.$$

The F test

The recipe for the F test:

- The null hypothesis is that $\beta_j = 0$ for all $j \in S$.
- If this is true, then the test statistic has an F distribution, with $p - |S|$ degrees of freedom in the numerator, and $n - p$ degrees of freedom in the denominator.
- We can use the F distribution to determine how unlikely our observed value of the test statistic is, if the null hypothesis were true.

Under the null we expect $F \approx 1$. Large values of F suggest we can reject the null.

The F test

The F test is a good example of why hypothesis testing is useful:

- We could implement the Wald test by just looking at the confidence interval for β_j.
- The same is not true for the F test: we can't determine whether we should reject the null by just looking at individual confidence intervals for each β_j.
- The F test is a succinct way to summarize our level of uncertainty *about multiple coefficients at once*.

"The" F test

Statistical software such as R does all the work for you.

First, note that regression output always includes information on "the" F statistic, e.g.:

```
Call:
lm(formula = Ozone ~ 1 + Solar.R + Wind + Temp, data = airqualit
...
F-statistic: 54.83 on 3 and 107 DF,  p-value: < 2.2e-16
```

This is always the F test against *the null that all coefficients (except the intercept) are zero*. What does rejecting this null mean? What is the alternative?

F tests of one model against another

More generally, you can use R to run an F test of one model against another:

```
> anova(fm_small, fm_big)
Analysis of Variance Table

Model 1: Ozone ~ 1 + Temp
Model 2: Ozone ~ 1 + Temp + Solar.R + Wind
  Res.Df   RSS Df Sum of Sq     F   Pr(>F)
1    109 62367
2    107 48003  2     14365 16.01 8.27e-07 ***
...
```

Caution

A word of warning

Used correctly, hypothesis tests are powerful tools to quantify your uncertainty. However, they can also be easily misused.

For example, suppose that you test and compare many models by repeatedly using F tests. What might go wrong?